DOCUMENT RESUME

ED 148 039 EC 103 068

AUTHOR	Heath, Robert W.
TITLE	The Application of Scaling Methodology to the Evaluation of Special Education Programs. Final Report.
INSTITUTION	NOMOS Inst., Berkeley, Calif.
SPONS AGENCY	California State Dept. of Education, Sacramento. Div. of Special Education.
PUB DATE	Aug 76
NOTE	42p.; State grant reference No. 75-151
EDRS PRICE	MF-$0.83 HC-$2.06 Plus Postage.
DESCRIPTORS	Comparative Analysis; Educable Mentally Handicapped; Elementary Education; Emotionally Disturbed; Evaluation Criteria; *Evaluation Methods; *Handicapped Children; Learning Disabilities; *Program Evaluation; Speech Handicapped

ABSTRACT

Described is a study in which a method of program evaluation was field tested on three types of elementary special education programs: speech impaired, educable mentally retarded, and educationally handicapped. Discussed is the development of a method which included the following attributes: direct observation; practicality; consideration of characteristics judged important by parents, pupils, and educators; and production of results which would be useful to improve the program. Statistical data is presented on pair-comparison of programs according to five attributes: individualization of instruction, global quality, pupil progress, pupil identification and evaluation, and program management. It is explained that the results of these pair-comparisons were analyzed and the ranking of programs was determined. Among conclusions listed is that the procedures identified and operationally defined characteristics of special programs important to those persons involved. (CL)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

NOMOS INSTITUTE 2372 ELLSWORTH BERKELEY, CALIFORNIA 94704 (415) 843-1107

ED148039

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

Final Report

Grant Ref. No. 75-151

The Application of Scaling Methodology to the Evaluation of Special Education Programs

August, 1976

EC103068

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
Stuart Greenfeld
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) AND USERS OF THE ERIC SYSTEM."

Contents

I.	Introduction	1
II.	Method	4
III.	Results	11
IV.	Conclusions	27
	Footnotes	30
	References	31
	Appendix A	32
	Appendix B	39

I. Introduction

How should educational programs be evaluated? How should the dimensions of assessment be selected? Where should one look for data on these dimensions? How can we know that the assessment is reliable?

This project attempted to work through these questions with a fresh point-of-view and without allegiance to any prior theories or techniques.

Rather than inferring the quality of a program from pupil test scores, expenditures per pupil, or attendance data, it seems reasonable to observe the relevant dimensions of program quality first hand. Rather than base program evaluation on variables selected on the basis of convenience to the evaluator, it seems better to ask parents, pupils and educators to identify those attributes of a program that are important to them. Rather than attempting to estimate the reliability of program evaluation by analyzing test scores (that is, pupil measures) we should attempt to analyze the reliability of the scores of programs based on those dimensions of importance to the constituents of the program.

The purpose of this research was to try a new method of educational program evaluation consistent with these premises. The method was designed to meet these requirements:

1. Programs should be evaluated on attributes judged to be important by parents, pupils, and educators.

2. These attributes of educational programs should be observed as directly as possible. Indirect inference about these attributes should be avoided.

3. The assessment of programs on these attributes should yield

reliable discrimination (rank-order) among the programs.

4. The method of program evaluation should be practical. It should not require excessive time or expense.

5. The method of program evaluation should yield results that are useful in the improvement of existing programs or in the planning of future programs.

In this study, we had the invaluable cooperation of the office of the Superintendent, of the Contra Costa County Schools.[1] Contra Costa County has 16 elementary school districts. This study field-tested the experimental method of program evaluation on three types of special education programs: Speech, Educable Mentally Retarded, and Educationally Handicapped.

Parents, educators, and pupils were interviewed to identify those attributes of these three types of program that were important to them. The results (written reports of the interviews) were subjected to content analysis (a social-service technique for quantifying interview data). Using the interviewers impressions, written records of the interviews, and the results of the content analysis, a modified Delphi technique (a prediction-refinement technique originally developed at the Rand Corporation) was employed to consolidate and operationally-define the attributes identified in the interviews. Observers were then trained to observe these attributes of school program, on-site, and to record their observations. The observer then compared each program (within special education category) to every other program and decided which was "greater" on each attribute (pair-comparison). The results of these pair-comparisons by the observers were subjected to statistical analysis and the reliability of the ranking of programs was determined.

This study was one of methodological research. It was not an application of conventional methods to a conventional problem. We had no reason to expect that everything we tried would succeed — and it didn't. In general however, the method worked. We know how to evaluate educational programs better than we did before this study. And we know how to avoid or prevent some of the problems we encountered in this trial-run.

II. Method

A. Determination of Dimensions

The first step in establishing criteria was to interview those people directly involved in special education programs. The observers interviewed special education teachers, students, and parents; special education administrators at the state, county, and district level; and faculty at the University of California. None of these interviews were conducted in Contra Costa County. With the exception of students, all persons interviewed were asked what aspects of special education they considered important. Students were asked how they would choose what program they would enter if they had a choice.

The written summaries of these "priority" interviews were content-analyzed to establish categories of program characteristics and the frequency of their mention. The results of the content analysis are presented in Table 1.

Insert Table 1 about here

A modified Delphi technique was employed to define and elaborate on these program characteristics (or dimensions) and to train the observers for site visits.

Based on their experiences in the interviews and the content analysis, each observer listed 10 attributes of each program that was considered indicative of high quality. For example, one attribute listed for EH was "there should be a low student-teacher ratio." For Speech, one of the attributes stipulated "the classroom should be soundproof and free of distractions." Each observer reviewed the attributes nominated by the

Table 1

Category and Frequency of Responses by Participants in Special Education Programs (n = 80)

	Program		
Category	EH	EMR	SPEECH
Effective Use of Staff Expertise and Coordination	59	44	25
Quality and Style of Teacher	72	54	58
Quality of Educational Program	80	85	56
Pupil Satisfaction	42	28	23
Administration	46	39	14
Identification	22	17	16
Assess/Amelioration	26	25	18
Parent Contact and Involvement	32	30	31
Integration	35	40	8

other observers and changed their own list of attributes where omission, inconsistency, or vagueness was apparent. Each observer then ranked their attributes on their importance to program quality, and the rankings of all the observers were compared for each program. Attributes commonly highly-ranked were designated general attributes, and specific attributes were collapsed into appropriate general attributes. Rankings of these general attributes were then performed by the observers and means and standard deviations calculated for each attribute in each program. The results of these calculations are presented in Table 2.

Insert Table 2 about here

The attributes in Table 2 were selected for use in site visits and scaling on the basis of the following criteria:

1. The average observer ranking (as presented in Table 2) must exceed 6.0.

2. It must be likely to vary more <u>between</u> programs than <u>within</u> programs.

3. It must be a <u>program</u> characteristic as opposed to a <u>pupil</u> or <u>teacher</u> characteristic.

4. It must be observable.

For the EH Program, Integration and Parent Involvement did not meet the first criterion, while Teacher Quality and Pupil Satisfaction did not meet the other three criteria; for the EMR Program, Individualization did not meet the first criterion, while Teacher Quality and Pupil Satisfaction did not meet the other three criteria; and for the Speech Program, Parent Involvement and Individualization did not meet the first criterion, and

Table 2

Means and Standard Deviations of Observer Rankings on Attributes, by Program

	EH		EMR		SPEECH	
Attribute	$\bar{X}$	S.D.	$\bar{X}$	S.D.	$\bar{X}$	S.D.
Pupil Progress	2.8	2.9	1.8	1.3	3.0	2.3
Pupil Satisfaction	3.0	2.3				
Teacher Quality and Personal Characteristics	4.0	1.2	2.8	1.1	2.6	1.1
Program Management	4.6	1.7	5.2	1.9	4.0	1.0
Pupil Identification and Evaluation	5.2	3.7	2.8	1.3	1.6	.55
Individualization	5.4	2.0	6.0	1.0		
Integration	6.6	1.5	4.0	2.1		
Parent Involvement	6.4	1.3				
Pupil Morale					4.4	1.1
Facilities, Equipment and Materials					5.4	.90

Teacher Quality, Pupil Satisfaction, and Facilities, Equipment, and Materials did not meet the other three criteria.

When the final list of attributes was established, each observer wrote definitions of each attribute and suggestions for sources of information on that attribute. The definitions and suggestions were circulated among observers and consensus was reached for each attribute. A site-visit packet was compiled consisting of definitions of each attribute (or dimension) and sources of information, including example questions applicable to each attribute and program (see Appendix A).

B. Site Visits

Site visits were conducted during November, 1975. The schedule was designed for each observer to contact each of the 15 districts[2] that agreed to be involved in the project, and for each of the districts to be visited at varying times during the school day. In each district, two observers contacted the district office and interviewed the Special Education Administrator, the School Psychologist, and requested specific information from program files. Two observers sat in on as many Special Education classes as possible and interviewed the teachers; and then another observer contacted and interviewed parents of Special Education children. Each of the observers performed an equal number of administrative interviews, classroom observations and teacher interviews, and parent interviews.

C. Scaling

As each site visit was completed, each observer recorded the results on Synopsis Cards, one card for each attribute in each program in each district. Without discussion, each observer then compared each district on each attribute using the pair-comparison technique. After the comparisons

had been completed, the observers performed a second round of comparisons without referring to their first round, and then reconciled differences between the two sets of results. This process yielded five comparison matrices for each program attribute (see Appendix B).

D. Data Analysis

For purposes of quality control, the comparison matrices were checked for inappropriate entries (for example, a "2" in the cell representing the comparison between District 4 and District 5), and for unequal marginal totals. Errors were referred to the observers for correction. One observer's matrices contained errors in more than half of the cases; these data were eliminated from further analysis. In addition, after the data were key-punched, they were verified, and inconsistencies in the matrices were resolved.

Before the data were key-punched and computer-analyzed, a trial pair-comparison scaling on one attribute was performed by hand. The calculations involved in obtaining scale values for four observers of 15 districts was approximately four hours, however calculations pertaining to the reliability of scaling took considerably longer than four hours, suggesting that hand calculation of reliability would be prohibitively time-consuming.

The decision to computer-analyze the data was made after the trial scaling had been completed and the level of effort involved in the procedure had been estimated.

E. Pair Comparison Scaling

In pair-comparison scaling, all stimuli under examination are presented to the observers in all possible pairs, and in each case the

observers judge which of the pair possesses greater value on a specified dimension or attribute. The data from these judgements are summarized in terms of the proportion of observers who designated each stimulus as "greater" on that attribute than each of the other stimuli. The arithmetic mean of the proportions for each stimulus is then calculated, yielding a scale value that indicates the rank of the district on the dimension under examination and its distance from the other districts on that dimension.

III. Results

The procedures just described began with a rational selection of dimensions on which to evaluate three kinds of special education programs and ended with dichotomous comparison of programs on these dimensions by four trained observers. These comparisons are summarized in the program X program matrices given in Appendix B. "Scaling" is the statistical analysis of the matrices to locate each program in rank-order on the dimensions and also to estimate the relative distance between each of the programs in this rank-order. That is, scaling produces an interval scale on each dimension.

All measurements, other than those on physical scales (weight, distance, etc.), yields ordinal scales. Intelligence tests, achievement tests, and so forth yield ordinal scale data that is treated as if it were interval scale data. In general, the reliability of educational and psychological measures is expressed (and statistically defined) in terms of how closely the observed data approximate a perfect interval scale. A perfect interval scale is perfectly internally-consistent. Thus, reliability becomes operationally defined as the ratio of how internally consistent a set of data is to perfect internal consistency (an interval scale without error).

Three different scaling methods were used to analyze the data.[3] The three methods were Guttman Scalogram Analysis, the traditional Thurstone Pair-Comparison Analysis, and SAATY - a method developed from recent work in multi-dimensional scaling. These three methods came from different historical traditions in measurement, involving differing sets of assumptions and take differing computational paths.[4]

If the method of program evaluation we field-tested provides reliable results, these results should look reliable across differing methods of analysis. We were interested to learn if these three methods give consistent results.

Consequently, the scale value of each program on each dimension was calculated by each of the three scaling techniques. A reliability index for each was calculated. Finally, the scale values produced by each of the three scaling techniques were correlated with the scale values produced by the other two scaling techniques.

The results of these analyses are given in Tables 3 through 16.

Insert Tables 3-16 about here

Table 3

Scale Values, Scalability Indexes, and Intercorrelations

Produced by Three Scaling Techniques

Type of Program: Educationally Handicapped

Attribute: Individualization of Instruction

Program (District)	Scale Values		
	Thurstone	Guttman	SAATY
1	.088	39	.335
2	1.248	4	.271
3	3.025	-48	.053
4	0.0	46	.385
5	2.578	-36	.095
6	1.338	-4	.240
7	1.155	20	.281
8	1.333	2	.265
9	1.515	-4	.244
10	2.847	-43	.059
11	2.668	-44	.068
12	.268	32	.350
13	1.425	-12	.210
14	.357	26	.333
15	.178	24	.319
Scalability Index	.72	.89	.90

Correlation of Scale Values Produced by the Three Techniques

	Thurstone	Guttman	SAATY
Thurstone			
Guttman	.982		
SAATY	.980	.985	

Table 4

Scale Values, Scalability Indexes, and Intercorrelations

Produced by Three Scaling Techniques

Type of Program: Educationally Handicapped

Attribute: Global Quality

Program (District)	Scale Values		
	Thurstone	Guttman	SAATY
1	.620	38	.333
2	.888	24	.309
3	3.378	-56	.037
4	1.691	4	.227
5	1.602	-4	.249
6	1.068	8	.285
7	.177	42	.379
8	2.578	-28	.134
9	2.578	-32	.123
10	1.333	2	.260
11	2.491	-32	.119
12	.978	8	.263
13	2.668	-40	.104
14	.622	28	.344
15	0.0	38	.375
Scalability Index	.72	.89	.90

Correlation of Scale Values Produced by the Three Techniques

	Thurstone	Guttman	SAATY
Thurstone			
Guttman	.983		
SAATY	.992	.989	

Table 5

Scale Values, Scalability Indexes, and Intercorrelations

Produced by Three Scaling Techniques

Type of Program: Educationally Handicapped

Attribute: Pupil Progress

Program	Scale Values		
(District)	Thurstone	Guttman	SAATY
1	1.868	12	.261
2	2.313	-6	.214
3	3.380	-54	.046
4	.710	38	.366
5	1.426	-6	.242
6	2.136	-14	.197
7	1.068	24	.337
8	2.847	-38	.129
9	1.690	-6	.217
10	.980	6	.253
11	.891	8	.258
12	.977	26	.327
13	3.113	-46	.072
14	0.0	34	.358
15	.623	22	.323
Scalability Index	.61	.80	.90

Correlation of Scale Values Produced by the Three Techniques

	Thurstone	Guttman	SAATY
Thurstone			
Guttman	.933		
SAATY	.939	.993	

Table 6

Scale Values, Scalability Indexes, and Intercorrelations

Produced by Three Scaling Techniques

Type of Program: Educationally Handicapped

Attribute: Pupil Identification and Evaluation

Program (District)	Scale Values		
	Thurstone	Guttman	SAATY
1	1.512	20	.312
2	1.333	22	.325
3	3.203	-38	.081
4	2.400	-6	.192
5	.890	26	.374
6	.800	26	.351
7	0.0	42	.440
8	2.580	-18	.168
9	3.025	-36	.095
10	1.873	-26	.096
11	.622	20	.322
12	1.868	-16	.128
13	1.958	-16	.137
14	1.515	-16	.128
15	.445	16	.309
Scalability Index	.51	.76	.89

Correlation of Scale Values Produced by the Three Techniques

	Thurstone	Guttman	SAATY
Thurstone			
Guttman	.890		
SAATY	.851	.987	

Table 7

Scale Values, Scalability Indexes, and Intercorrelations

Produced by Three Scaling Techniques

Type of Program: Educationally Handicapped

Attribute: Program Management

Program	Scale Values		
(District)	Thurstone	Guttman	SAATY
1	0.0	46	.385
2	.803	22	.319
3	3.115	-48	.051
4	2.760	-42	.067
5	1.690	2	.221
6	1.336	2	.272
7	.090	42	.380
8	2.403	-30	.093
9	2.315	-28	.123
10	1.602	-4	.171
11	2.047	-18	.174
12	.448	20	.348
13	2.226	-30	.118
14	.447	26	.322
15	.092	40	.392
Scalability Index	.75	.88	.91

Correlation of Scale Values Produced by the Three Techniques

	Thurstone	Guttman	SAATY
Thurstone			
Guttman	.990		
SAATY	.987	.985	

Table 8

Scale Values, Scalability Indexes, and Intercorrelations

Produced by Three Scaling Techniques

Type of Program: Educable Mentally Retarded

Attribute: Integration

Program	Scale Values		
(District)	Thurstone	Guttman	SAATY
1	.667	16	.619
5	2.667	-8	.174
8	0.0	16	.619
10	1.554	-2	.320
13	3.112	-18	.088
14	2.0	-4	.306
Scalability Index	.77	.88	.94

Correlation of Scale Values Produced by the Three Techniques

	Thurstone	Guttman	SAATY
Thurstone			
Guttman	.970		
SAATY	.976	.993	

Table 9

Scale Values, Scalability Indexes, and Intercorrelations

Produced by Three Scaling Techniques

Type of Program: Educable Mentally Retarded

Attribute: Pupil Progress

Program (District)	Scale Values		
	Thurstone	Guttman	SAATY
1	1.112	6	.425
5	.671	2	.390
8	1.562	-6	.344
10	1.112	2	.408
13	2.891	-20	.095
14	0.0	16	.610
Scalability Index	.63	.82	.93

Correlation of Scale Values Produced by the Three Techniques

	Thurstone	Guttman	SAATY
Thurstone			
Guttman	.966		
SAATY	.968	.986	

Table 10

Scale Values, Scalability Indexes, and Intercorrelations

Produced by Three Scaling Techniques

Type of Program: Educable Mentally Retarded

Attribute: Pupil Identification and Evaluation

Program	Scale Values		
(District)	Thurstone	Guttman	SAATY
1	.225	12	.563
5	1.112	2	.380
8	2.004	-10	.174
10	1.558	-4	.334
13	2.450	-16	.102
14	0.0	16	.621
Scalability Index	.77	.88	.88

Correlation of Scale Values Produced by the Three Techniques

	Thurstone	Guttman	SAATY
Thurstone			
Guttman	.999		
SAATY	.994	.995	

Table 11

Scale Values, Scalability Indexes, and Intercorrelations

Produced by Three Scaling Techniques

Type of Program: Educable Mentally Retarded

Attribute: Program Management

Program	Scale Values		
(District)	Thurstone	Guttman	SAATY
1	0.0	20	.743
5	2.891	-12	.131
8	2.888	-10	.174
10	1.554	6	.488
13	1.779	-6	.174
14	.888	2	.364
Scalability Index	.63	.82	.92

Correlation of Scale Values Produced by the Three Techniques

	Thurstone	Guttman	SAATY
Thurstone			
Guttman	.930		
SAATY	.879	.991	

Table 12

Scale Values, Scalability Indexes, and Intercorrelations

Produced by Three Scaling Techniques

Type of Program: Educable Mentally Retarded

Attribute: Global Quality

Program (District)	Scale Values		
	Thurstone	Guttman	SAATY
1	0.0	20	.738
5	2.000	4	.467
8	2.667	-12	.161
10	1.775	0.0	.253
13	2.671	-14	.117
14	.888	2	.365
Scalability Index	.67	.83	.91

Correlation of Scale Values Produced by the Three Techniques

	Thurstone	Guttman	SAATY
Thurstone			
Guttman	.921		
SAATY	.873	.974	

Table 13

Scale Values, Scalability Indexes, and Intercorrelations

Produced by Three Scaling Techniques

Type of Program: Speech

Attribute: Pupil Progress

Program (District)	Scale Values		
	Thurstone	Guttman	SAATY
1	.761	24	.336
2	.762	22	.407
3	.472	28	.331
5	1.812	-10	.185
6	.565	28	.258
7	1.523	-2	.235
8	2.668	-40	.079
9	0.0	28	.412
10	1.143	4	.284
11	2.575	-36	.077
12	1.905	-18	.174
13	2.287	-34	.095
14	.762	2	.239
15	.764	4	.313
Scalability Index	.60	.79	.87

Correlation of Scale Values Produced by the Three Techniques

	Thurstone	Guttman	SAATY
Thurstone			
Guttman	.958		
SAATY	.927	.926	

Table 14

Scale Values, Scalability Indexes, and Intercorrelations

Produced by Three Scaling Techniques

Type of Program: Speech

Attribute: Pupil Identification and Evaluation

Program	Scale Values		
(District)	Thurstone	Guttman	SAATY
1	2.862	-34	.081
2	1.145	26	.363
3	2.478	-16	.133
5	1.334	6	.276
6	.764	16	.310
7	.681	16	.355
8	2.193	-24	.117
9	2.480	-25	.109
10	.368	34	.379
11	1.145	0.0	.221
12	2.287	-25	.115
13	2.385	-36	.080
14	.573	26	.369
15	0.0	36	.414
Scalability Index	.62	.80	.91

Correlation of Scale Values Produced by the Three Techniques

	Thurstone	Guttman	SAATY
Thurstone			
Guttman	.957		
SAATY	.962	.991	

Table 15

Scale Values, Scalability Indexes, and Intercorrelations

Produced by Three Scaling Techniques

Type of Program: Speech

Attribute: Program Management

Program	Scale Values		
(District)	Thurstone	Guttman	SAATY
1	2.003	-6	.229
2	1.146	6	.299
3	3.336	-50	.043
5	1.334	20	.318
6	1.830	-9	.179
7	1.048	24	.312
8	2.668	-29	.162
9	1.923	-13	.226
10	.096	46	.416
11	3.032	-41	.067
12	.668	18	.316
13	2.463	-27	.127
14	.477	23	.331
15	0.0	38	.382
Scalability Index	.73	.87	.90

Correlation of Scale Values Produced by the Three Techniques

	Thurstone	Guttman	SAATY
Thurstone			
Guttman	.977		
SAATY	.970	.985	

Table 16

Scale Values, Scalability Indexes, and Intercorrelations

Produced by Three Scaling Techniques

Type of Program: Speech

Attribute: Global Quality

Program	Scale Values		
(District)	Thurstone	Guttman	SAATY
1	1.812	0.0	.253
2	1.336	10	.290
3	3.430	-44	.062
5	1.336	6	.283
6	.575	26	.375
7	1.046	18	.282
8	2.955	-44	.059
9	2.003	-10	.237
10	.380	40	.389
11	2.194	-20	.185
12	1.336	-6	.211
13	2.859	-40	.077
14	.761	26	.315
15	0.0	38	.391
Scalability Index	.70	.85	.91

Correlation of Scale Values Produced by the Three Techniques

	Thurstone	Guttman	SAATY
Thurstone			
Guttman	.980		
SAATY	.966	.985	

IV. Conclusions

Study of these tables supports the following conclusions:

A. Some attributes (dimensions) of the program are more scalable than others, at least by the methods used in this study. We can categorize the results as yielding "adequate" or "inadequate" internal consistency.

Insert Table 17 about here

Of the 14 dimensions tested for internal consistency (scalability) 10 produce data of "adequate" reliability and four did not. It does not seem to be the attribute in general that is inherently scalable or not scalable. An attribute may be scalable in one type of program but not in another. This seems a very important fact that is frequently overlooked in the design of educational program evaluation.

B. Though the nature of the indexes produced by the three scaling methods are different, the results are generally consistent. The correlations among the scale values are high. Correlation coefficients of this magnitude cannot be obtained unless both the variables being correlated have substantial internal consistency. Unreliable measures just don't correlate with anything. The differences in the scalability coefficients produced is primarily a function of the coefficient's sensitivity to a small number of judges. The Thurstone method assumes a large number while the SAATY is quite "robust" in this regard.

C. The procedures used here successfully identified and operationally defined those attributes of special education programs of importance to

Table 17

Scalability of Program Attributes

Program/Attribute	Adequate Scalability	Inadequate Scalability
Educationally Handicapped		
Individualization	X	
Global Quality	X	
Pupil Progress		X
Pupil Identification and Evaluation		X
Program Management	X	
Educable Mentally Retarded		
Individualization	X	
Global Quality		X
Pupil Progress	X	
Pupil Identification and Evaluation	X	
Program Management	X	
Speech		
Global Quality	X	
Pupil Progress		X
Pupil Identification and Evaluation	X	
Program Management	X	
Total	10	4

the people they serve. Other procedures might have done as well, or better, but at least we know that variables other than achievement test scores, per/pupil expenditures, and attendance are feasible. The importance of this is that the variables were used because they were of expressed concern to those involved -- they were not selected because they were "available" or convenient to measure. The validity of these attributes has a rational basis.

D. Though a cost-effectiveness study was not planned or conducted, it seems clear that this type of program evaluation is economically and administratively efficient. Three different types of programs in 15 school districts were evaluated in a few weeks time at minimal cost (non-research costs were probably less than $15,000). The experience gained in this research study should make it easier and more economical for similar applications of scale analysis to the evaluation of educational programs in the future.

Footnotes

[1]In particular we are grateful to Harry Larsen for his generous assistance and valuable advice.

[2]One of the 16 elementary districts in Contra Costa County chose not to cooperate with the study.

[3]Since no novel scaling methods were used, they will not be detailed here. The references cited document the scaling methods.

[4]We are indebted to Tom Reynolds who identified the computer program and ran our data for us at U.S.C.

References

Edwards, A. L. Techniques of Attitude Scale Construction. New York: Appleton-Century-Crofts, Inc., 1957.

Guilford, J. P. Psychometric Methods. New York: McGraw Hill, 1954.

Reynolds, T. J. Analysis of Dominance Matrices: Extraction of Unidimensional Orders within a Multidimensional Context. Unpublished doctoral dissertation, University of Southern California, 1976.

Appendix A: Definitions of Dimensions and Indicators of Program Quality

This Appendix contains materials used by observers during the site visits. For each program and dimension selected (according to the procedure detailed in the Methods section of this report), observers agreed both on a precise definition of program quality and on positive and negative indicators of program quality that could be observed during site visits. The definitions of program quality and examples of indicators of program quality follow.

EH: Pupil Identification and Reassessment

Definition of Dimension: All pupils needing EH services are identified early and accurately and are reassessed frequently by a variety of qualified personnel, as well as reliable, valid instruments and procedures.

Positive Indicators:

—Referrals come from a variety of sources (e.g., parents, teachers, principals, nurses, speech therapists, outside agencies, medical doctors).

—There is the presence of detailed records of screening, diagnostic, placement and reassessment procedures.

—The EH program conducts a needs assessment survey regularly and frequently.

—The EH program publicizes (disseminates information) to teachers, schools, community agencies, social welfare agencies, public health agencies.

Negative Indicator:

—There is evidence that pupils are receiving medication to control behavior. (Ritalin)

EH: Program Management

Definition of Dimension: The EH program efficiently plans for and coordinates an adequate distribution of program resources. Administration provides strong leadership and supervision of staff.

Positive Indicators:

—Program provides inservice training for EH teachers and regular teachers.

—Program provides money and release time for EH teachers to attend special seminars and continuing education programs in the field.

—Program provides for regular meetings between EH teachers and regular teachers.

—EH teachers feel they get adequate support and information from regular teachers.

Negative Indicator:

—The EH teacher spends an excessive amount of time traveling between sites.

EH: Individualization

Definition of Dimension: Services provided by the programs are selected and modified on the basis of a documented analysis of each child's individual needs.

Positive Indicators:

—The emotional and physical needs of each child are taken into account.

—The program is individualized for each pupil, academically, physically, and psychologically.

—The program provides records of pupil progress.

—There are individual instructional plans for each pupil.

Negative Indicator:

—All pupils have same materials, but working at their own pace.

EH: Pupil Progress

Definition of Dimension: Pupils served by the program show documentable, observable improvement (corresponding to their diagnosed needs) associated with the program's efforts.

Positive Indicators:

—There is a high rate of return to the regular program.

—There is a high rate of movement from special day classes into DIS (Designated Instructional Service).

—Average length of time in the EH program is from two to three years.

—Test Scores demonstrate improvement.

Negative Indicator:

—There is evidence that medication is used to control the pupils' behavior (Ritalin).

EMR: Pupil Progress

Definition of Dimension: Pupils served by the program show documentable, observable improvement (corresponding to their diagnosed needs) associated with the program's efforts.

Positive Indicators:

—Test scores demonstrate improvement (academic, physical, and psychological).

—There is a high rate of return to the regular, or EH program.

—Parents feel that the child has made improvement.

—Pupils demonstrate self-sufficiency appropriate to their ability.

EMR: Integration

Definition of Dimension: EMR pupils are optimally integrated into the regular school program and are accepted by the regular pupils and staff.

Positive Indicators:

—EMR pupils do not express fear of integrated situations.

—EMR pupils feel welcome in the school.

—EMR special day classrooms adjoin regular classrooms.

—EMR pupils share common facilities with regular pupils, such as rest rooms, lunchrooms, the playground, and library.

EMR: Pupil Identification and Reassessment

Definition of Dimension: All pupils needing EMR services are identified early and accurately and are reassessed frequently by a variety of qualified personnel, as well as reliable, valid instruments and procedures.

Positive Indicators:

—Referrals come from a variety of sources (e.g., parents, teachers, principals, nurses, speech therapists, outside agencies, medical doctors).

—There is the presence of detailed records of screening, diagnostic, placement and reassessment procedures.

—The EMR program conducts a needs assessment survey regularly and frequently.

—The EMR program publicizes (disseminates information) to teachers, schools, community agencies, social welfare agencies, public health agencies.

EMR: Program Management

Definition of Dimension: The EMR program efficiently plans for and coordinates an adequate distribution of program resources. Administration provides strong leadership and supervision of staff.

Positive Indicators:

—Program provides in-service training for EMR teachers and regular teachers.

—Program provides money and release time for EMR teacher to attend special seminars and continuing education programs in the field.

—Program provides for regular meetings between EMR teachers and regular teachers.

—EMR teachers feel they get adequate support from regular teachers.

Speech: Pupil Progress

Definition of Dimension: Pupils served by the program show documentable, observable improvement (corresponding to their diagnosed needs) associated with the program's efforts.

Positive Indicators:

—There are records that show pupil improvement, aggregate as well as individual.

—There is a high rate of movement out of the speech program.

—Pupils speak more often in regular classroom.

—Pupils speak more often outside of speech class.

Speech: Pupil Identification and Reassessment

Definition of Dimension: All pupils served by the program show documentable/observable improvement (corresponding to their diagnosed needs) associated with the program's efforts.

Positive Indicators:

—Every child is screened upon entering the district.

—Each child in school is screened at the beginning of the school year.

—There are a variety of sources of referrals into the speech program (e.g., parents, teachers, principals, nurses, outside agencies, medical doctors, etc.).

—There is the presence of detailed records of screening, diagnostic, placement and reassessment procedures.

Speech: Program Management

Definition of Dimension: The speech program efficiently plans for and coordinates an adequate distribution of program resources. Administration provides strong leadership and supervision of staff.

Positive Indicators:

—Program provides in-service training for speech teachers as well as regular teachers.

—Program provides money and release time for therapist to attend special seminars and upgrading programs in speech.

—Program provides for regular meetings between speech and regular staff.

—Speech therapists in the district meet regularly.

Negative Indicator:

—The therapist does spend an excessive amount of time in travel between sites.

Appendix B: Comparison Matrices

This Appendix contains examples of the comparison matrices used by judges to compare programs in each district with those in each other district. Judges wrote the number of the district they judged to have greater value on the given attribute in the cell corresponding to each pair of districts. Each comparison matrix was checked for internal consistency.